

1. Datos

- Cargar datos de R (*.RData *.rda): *Datos / Cargar conjunto de datos...* Seleccionar el fichero y verificar que se ha cargado adecuadamente (en la parte de abajo dice cuantas filas y columnas hay y si pinchamos en la opción *Visualizar conjunto de datos* nos muestra los datos en una ventana)
- Cargar datos de excell: *Datos / Importar datos... / Desde un archivo de excell...* decir si los nombres de las variables están en la cabecera del fichero excell y cargar el fichero *.xlsx.
- Comparar datos. Si queremos comparar dos variables que están en el mismo conjunto de datos, lo más cómodo es apilarlos, es decir, colocar una variable debajo de la otra y el R-Commander nos genera una segunda variable (factor) que especifica a qué variable original pertenecía cada dato. *Datos / Conjunto de datos activo / Apilar conjunto de datos activo*
- Convertir en factor. Si una variable esta codificada con números, el R-Commander lo entiende como una variable cuantitativa, para convertirla en cualitativa o factor: *Datos / Modificar variables del conjunto de datos activo / Convertir variable numérica en factor...* Una vez que sea cualitativa o factor no se puede recuperar la cuantitativa, por lo que si interesa conservarla hay que cambiar el nombre.
- Filtrar datos. Si queremos analizar solo una parte de los datos. *Datos / Conjunto de datos activo / Filtrar el conjunto de datos activo*. Si la variable de filtrado es cuantitativa, por ejemplo, queremos las observaciones con la variable edad menor de 35 años, en la *Expresión de la selección* se escribe, *edad < 35*. Si la variable de filtrado es un factor, por ejemplo, en los datos de coleopteros queremos separar los machos, se escribe *sexo == 0*. En ambos casos hay que cambiar el nombre del fichero si no queremos perder la base de datos original.
- Borrar filas. Anotar qué filas queremos quitar, *Datos / Conjunto de datos activo / Borrar fila(s) del conjunto de datos activo*. Escribir las filas que queremos quitar, sin separador o separadas por comas y dar nombre al nuevo conjunto de datos si no queremos perder la base de datos original.
- Transformar datos. *Datos / Modificar variables del conjunto de datos activo / Calcular una nueva variable....* Nos pide el nombre de la nueva variable y la expresión a calcular (las habituales $1/\text{variable}$; variable^2 ; $\text{sqrt}(\text{variable})$ y $\log(\text{variable})$). También está la opción de tipificar variables (resta la media y divide por la desviación típica): *Datos / Conjunto de datos activo / Tipificar variable*.

2. Resúmenes numéricos

- En la opción *Estadísticos / Resúmenes / Conjunto de datos activo*, se obtiene: para las variables numéricas, máximos y mínimos, el primer y tercer cuartil (Q_1 y Q_3), la mediana, la media, así como el número de valores que faltan. Para las variables cualitativas: el número de observaciones de cada nivel del factor.
- En la opción *Estadísticos / Resúmenes / Resúmenes numéricos...*, sólo para variables numéricas, da opción para calcular muchos estadísticos. La opción de *Resumir por grupos...*, permite obtener los estadísticos resumen para cada valor o nivel de un factor. La pestaña de *Estadísticos* permite cambiar los estadísticos que se calculan por defecto.
- Tabla de frecuencias. Variable cualitativas o factores: en la opción *Estadísticos / Resúmenes / Distribución de frecuencias...*, sólo para las variables cualitativas o factores. Obtenemos la tabla de frecuencias absolutas y relativas. Para variables cuantitativas hay que introducir el comando *hist(archivo\$variables, breaks=numero-de-clases, plot=FALSE)*
- Tablas de doble entrada. *Estadísticos / Tabla de contingencia / Introducir y analizar una tabla de doble entrada*. Se ha de rellenar la tabla completa y da opción a sacar una de las marginales, ambas o ninguna.

3. Gráficos

- Histograma. *Gráficas / Histograma....* Sólo variables cuantitativas. Escoger una variable. Se puede analizar por grupos de un factor y cambiar el número de clases.
- Diagrama de cajas. *Gráficas / Diagrama de caja....* Sólo variables cuantitativas. Escoger una variable. Se puede analizar por grupos de un factor y cambiar la forma de identificación de los atípicos.
- Gráfica de barras. *Gráficas / Gráfica de barras....* Sólo variables cualitativas.
- Gráfica de sectores. *Gráficas / Gráfica de sectores....* Sólo variables cualitativas.

4. Probabilidad *Distribuciones / Modelos continuos / Modelos discretos*, una vez elegido el modelo de probabilidad con el que se desea trabajar el programa ofrece:

- Distribuciones continuas.
 - a) Cuantiles: valores de la inversa de la función de distribución del modelo.
 - b) Probabilidades: pide valor(es) de la variable, los parámetros de la distribución y permite cola izquierda o derecha.
 - c) Gráfica de la distribución: da la opción de la función de densidad o la de distribución.
 - d) Muestra de la distribución: para generar muestras aleatorias simples.

- Distribuciones discretas.
 - a) Cuantiles: valores de la inversa de la función de distribución del modelo.
 - b) Probabilidades acumuladas: pide valor(es) de la variable, los parámetros de la distribución y permite cola izquierda o derecha. Si es la cola de la izquierda, calcula $P(X \leq k)$, menor o igual, si es la cola de la derecha, calcula $P(X > k)$, mayor estricto.
 - c) Probabilidades: pide los parámetros de la distribución y saca la función de probabilidad.
 - d) Gráfica de la distribución: da la opción de la función de probabilidad o de la de distribución.
 - e) Muestra de la distribución: para generar muestras aleatorias simples.

5. Inferencia

- Paramétrica

Sólo con una muestra de datos, para la media muestral con varianza desconocida, hace intervalos de confianza y contrastes unilaterales o bilaterales a cualquier nivel. *Estadístico / Media / Test t para una muestra.*
- No paramétrica

Contrastes de normalidad, para $n \leq 50$ usamos Shapiro-Wilks: *Estadístico / Resúmenes / Test de normalidad Shapiro-Wilks.*

Contrastes de ajuste de forma para distribuciones continuas, para $n \geq 30$ usamos Kolmogorov- Smirnov: *ks.test(archivo\$variable, distribución, parámetros)*

Por ejemplo para la normal: *ks.test(archivo\$variable, "pnorm", mu, sigma)*
 para la exponencial *ks.test(archivo\$variable, "pexp", λ).*

6. Análisis de la Varianza

- Hay veces que los datos por en columnas por factor, en ese caso hay que apilar: *Datos \rightarrow Conjunto de datos activo \rightarrow apilar variables del conjunto de datos activo...*
- Representación gráfica de los datos: *Gráficas \rightarrow Diagrama de puntos y/o Diagrama de cajas...*
- En algunos casos se debe convertir una variable en factor: *Datos \rightarrow Conjunto de datos activo \rightarrow convertir una variable numérica en un factor...*
- Para obtener la tabla ADEVA: *Estadísticos \rightarrow Medias \rightarrow Anova de un factor.*
- En ese mismo cuadro de diálogo se puede marcar *comparaciones dos a dos de medias.*

7. Regresión Lineal

- Gráfico X-Y: *Gráficas \rightarrow Diagrama de dispersión y/o Gráfica XY.*
- Covarianza: se escribe la orden *cov(archivo\$var1,archivo\$var2)*
- Coeficiente de correlación: *Estadísticos \rightarrow Resúmenes \rightarrow Matriz de correlaciones.*
 Una vez obtenida la matriz de correlaciones si en la ventana de instrucciones cambiamos *cor* por *cov* se obtiene la matriz de varianzas-covarianzas.
- Para calcular los parámetros del modelo de regresión:
 - *Estadísticos \rightarrow Ajuste de modelos \rightarrow Regresión lineal*
 - Hay que marcar la *variable explicada* y la *variable explicativa* y permite introducir un nombre para el modelo.

8. Diagnósis

- Para obtener los gráficos de residuos: *Modelos \rightarrow Gráficas \rightarrow Gráficas básicas de diagnóstico*
- Generar la variable que recoge los residuos hay dos opciones
 - a) *Datos \rightarrow Modificar variables del conjunto de datos activo \rightarrow Calcular una nueva variable;* permite poner nombre a la nueva variable, y en expresión a calcular: *resid(nombre del modelo).*
 - b) Con el modelo estimado activo (está en azul arriba a la derecha) *Modelos \rightarrow Añadir Estadísticas de las observaciones de los datos* y seleccionar residuos.
 - Con la variable residuos se puede: aplicar test de normalidad Shapiro-Wilk, *Estadísticos \rightarrow Resúmenes \rightarrow Test de normalidad de Shapiro-Wilk...*; Histograma; Diagrama de caja... en *Gráficas...*
 - Test de normalidad de Kolmogorov-Smirnov *ks.test(archivo\$variable,pnorm,0, sd(archivo\$variable))*
- Homocedasticidad en Adeva:
 - Se pueden pasar los test de Bartlett y de Levene: *Estadísticos \rightarrow Varianzas \rightarrow Test de Bartlett... Test de Levene...*
 - Si hay muy pocos datos y todos los niveles tienen el mismo número de datos se mira el ratio entre la mayor y la menor varianza de entre los niveles y si es menor que 3 aceptaremos homocedasticidad.